LECTURE X. ANOTHER ABSTRACT NORMAL APPROXIMATION THEOREM

Here I shall describe an approach that is applicable to the study of sums of m-dependent random variables and to the stationary case with appropriate mixing conditions. In these cases it does not seem possible to construct an exchangeable pair (W,W') with the properties required in Theorem III.1 and in Lemmas I.3, I.4 and III.1, which lead up to that theorem. The method of the present lecture is that which I used in the easier part of my paper on normal approximation in the Proceedings of the Sixth Berkeley Symposium. However, the difficult problem of getting error bounds that are essentially sharp in order of magnitude will not be treated here.

The present treatment is somewhat sketchy. I omit a number of proofs that are completely analogous to proofs in the first three lectures. Also, even the relatively concrete Corollary 2 does not deal with the usual formulation of the problems considered here. Some notion of this can be obtained from my paper mentioned earlier. Finally, I should mention that, in Lecture XIV, I shall reformulate this approach in a way that will clarify its relation to the basic formalism of the first lecture through (I.33).

In order to avoid excessive repetition in stating the hypotheses of the lemmas and theorem of this lecture I shall formulate the

Basic assumption: A probability space $(\tilde{\Omega},\tilde{\mathcal{B}},\tilde{P})$ is given and $\mathcal{B}$ and $\mathcal{C}$ are sub-σ-algebras of $\tilde{\mathcal{B}}$. The real random variable G is $\tilde{\mathcal{B}}$-measurable and the random variable W^* is $\mathcal{C}$-measurable. Assuming

$$E|G| < \infty \tag{1}$$

we define

$$W = E^{\mathcal{B}}G. \tag{2}$$

Lemma 1: In addition to the basic assumption, suppose $f: R \to R$ is a bounded piecewise-continuous function. Then

$$E\, Wf(W) = E\, G(f(W) - f(W^*)) + E\, (E^{\mathcal{C}}G)f(W^*). \tag{3}$$

Proof:

$$\begin{aligned} E\, Wf(W) &= E\, (E^{\mathcal{B}}G)f(W) = E\, Gf(W) \\ &= E\, G(f(W)-f(W^*)) + E\, (E^{\mathcal{C}}G)f(W^*). \end{aligned} \tag{4}$$

Lemma 2: In addition to the basic assumption, suppose $h: R \to R$ is a bounded piecewise continuous function and let $\mathcal{N}h$ and $U_{\mathcal{N}}h$ be defined as in (II.2) and (II.4). Then

$$E\, h(W) = Nh + E\{f'(W) - G(f(W)-f(W^*))\} - E(E^{\mathcal{C}}G)f(W^*) \tag{5}$$

where $f = U_N h$.

The proof is omitted since it is analogous to the derivation of Lemma I.4 from Lemma I.3.

Lemma 3: In addition to the basic assumption, let $h: R \to R$ be a bounded continuous function with bounded piecewise continuous derivative h'. Then

$$\begin{aligned} E\, h(W) &= Nh + E\, f'(W)[1-G(W-W^*)] - E(E^{\mathcal{C}}G)f(W^*) \\ &+ \int_{-\infty}^{\infty} EG(z-W^*)[\mathcal{I}\{z \le W\} - \mathcal{I}\{z \le W^*\}]f''(z)dz. \end{aligned} \tag{6}$$

The proof is omitted since it is analogous to the derivation of Lemma III.1 from Lemma I.4.

Theorem 1: Under the hypotheses of Lemma 3

$$(7) \qquad |Eh(W)-Nh| \leq 2(\sup|h-Nh|)\sqrt{E[1-E^{\mathcal{B}}G(W-W^*)]^2}$$
$$+ \sqrt{\frac{\pi}{2}}\,(\sup|h-Nh|)E|E^{\mathcal{C}}G| + (\sup|h'|)E|G|(W-W^*)^2.$$

The proof is omitted since it is analogous to the derivation of the inequality (III.10) in Theorem III.1 from Lemma III.1 together with an application of the inequality (II.45) in Lemma II.3.

Corollary 1: Under the basic assumption,

$$(8) \qquad |P\{W \leq w_0\} - \Phi(w_0)|$$
$$\leq 2\sqrt{E[1-E^{\mathcal{B}}G(W-W^*)]^2} + \sqrt{\frac{\pi}{2}}\,E|E^{\mathcal{C}}G| + 2^{3/4}\pi^{-\frac{1}{4}}\sqrt{E|G|(W-W^*)^2}.$$

The proof is analogous to the derivation of (III.11) from (III.10).

Now let us look at a special case of Theorem 1, where we consider a sum of a large number of small random variables, most of which are nearly independent of most of the others. This includes the m-dependent case, the stationary case with an appropriate mixing condition, and some analogues of these with a multi-dimensional index set. As in the third lecture, these results are far from the best possible in that the bounds are not ordinarily sharp even in order of magnitude and the moment conditions are not the weakest possible.

Let $\mathcal{J}$ be a finite set and, for each $i \in \mathcal{J}$ let X_i be a real random variable such that

$$(9) \qquad E\,X_i = 0$$

and

$$(10) \qquad E\,X_i^4 < \infty.$$

Also let

$$(11) \qquad W = \sum_{i \in \mathcal{J}} X_i$$

and suppose that for each $i \in \mathcal{J}$ a subset $S_i \subset \mathcal{J}$ has been chosen in such a way that

(12) $$E \sum_{i\in\mathcal{J}} X_i \sum_{j\in S_i} X_j = 1.$$

A bound will be obtained in Corollary 2 for the error in the standard normal approximation to the distribution of W. Roughly speaking, this bound will be small if

(i) the number n of elements of the index set $\mathcal{J}$ is large,

(ii) most of the S_i are small,

(iii) for most of the $i \in \mathcal{J}$, X_i is nearly independent of the $\{X_j\}_{j\notin S_i}$,

(iv) each of the X_i contributes only a small proportion of the total variability of W.

In order to fit this problem into the framework of the Basic Assumption, I shall introduce an additional random variable I, uniformly distributed over the finite set $\mathcal{J}$ independent of the $\{X_j\}_{j\in\mathcal{J}}$. Let $\tilde{\mathcal{B}}$ be a σ-algebra in which the random variable I and the $\{X_j\}_{j\in\mathcal{J}}$ are measurable, let $\mathcal{B}$ be the sub-σ-algebra of $\tilde{\mathcal{B}}$ generated by the $\{X_j\}_{j\in\mathcal{J}}$ and, finally, let $\mathcal{C}$ be the σ-algebra generated by I and the $\{X_j\}_{j\notin S_I}$. A more precise description of $\mathcal{C}$ is that it consists of all $B \in \tilde{\mathcal{B}}$ having the property that, if $\omega, \omega' \in \tilde{\Omega}$ are such that

(13) $$I(\omega') = I(\omega)$$

and, for all $j \notin S_{I(\omega)}$,

(14) $$X_j(\omega') = X_j(\omega),$$

then

(15) $$\omega' \in B \Leftrightarrow \omega \in B.$$

It will ordinarily be simplest to choose $\tilde{\Omega}$ to consist of all (n+1)-uples $(i,\{x_j\}_{j\in\mathcal{J}})$ with $i \in \mathcal{J}$ and the $x_j \in R$ and to define the random variables I and

$\{X_j\}_{j\in\mathcal{J}}$ in the obvious way by

(16) $$I(i,\{x_{j'}\}_{j'\in\mathcal{J}}) = i$$

and

(17) $$X_j(i,\{x_{j'}\}_{j'\in\mathcal{J}}) = x_j.$$

However, it may sometimes be necessary to introduce a more complicated underlying space $\tilde{\Omega}$ in order to obtain better approximations by using additional randomization. The probability measure $\tilde{P}$ is chosen in such a way that these random variables have the distribution described earlier. Continuing to set up the correspondence between this situation and the Basic Assumption, let

(18) $$G = nX_I$$

so that

(19) $$E^{\mathcal{B}}G = \frac{1}{n}\sum_{i\in\mathcal{J}} nX_i = \sum_{i\in\mathcal{J}} X_i = W,$$

in agreement with (2) and (11). Finally let

(20) $$W^* = \sum_{j\notin S_I} X_j = W - \sum_{j\in S_I} X_j.$$

Of course we shall ordinarily choose the $\{S_i\}_{i\in\mathcal{J}}$ in such a way that $i \in S_i$.

Now let us evaluate the expectations occurring on the right hand side of (7) and (8) in the special situation considered here. We have

(21) $$E|E^{\mathcal{C}}G| = E\sum_{i\in\mathcal{J}} |E^{\{X_j\}_{j\notin S_i}}X_i|$$

and

(22) $$E|G|(W-W^*)^2 = E\sum_{i\in\mathcal{J}} |X_i|(\sum_{j\in S_i} X_j)^2,$$

and finally, with

(23) $$\sigma_{ij} = EX_iX_j,$$

also

$$(24) \qquad E[1-E^{\mathcal{B}}G(W-W^*)]^2 = E[\sum_{i\in\mathcal{J}} \sum_{j\in S_i} (X_iX_j-\sigma_{ij})]^2.$$

Let us summarize these computations in a corollary, which is now an immediate consequence of Theorem 1 and Corollary 1:

Corollary 2: Let $\mathcal{J}$ be a finite set and, for each $i \in \mathcal{J}$, X_i a real random variable and S_i a subset of $\mathcal{J}$ such that (9), (10) and (12) hold and let W be defined by (11). Then, for all real w_0,

$$(25) \qquad |P\{W\leq w_0\}-\Phi(w_0)| \leq 2\sqrt{E[\sum_{i\in\mathcal{J}} \sum_{j\in S_i} (X_iX_j-\sigma_{ij})]^2}$$

$$+ \sqrt{\frac{\pi}{2}}\, E\sum_{i\in\mathcal{J}} |E^{\{X_j\}_{j\notin S_i}} X_i| + 2^{3/4}\pi^{-\frac{1}{4}}\sqrt{E\sum_{i\in\mathcal{J}} |X_i|(\sum_{j\in S_i} X_j)^2},$$

and, for any bounded, continuous, and piecewise continuously differentiable function h: $R \to R$,

$$(26) \qquad |Eh(W)-Nh| \leq 2\sup|h-Nh| \sqrt{E[\sum_{i\in\mathcal{J}} \sum_{j\in S_i} (X_iX_j-\sigma_{ij})]^2}$$

$$+ \sqrt{\frac{\pi}{2}} \sup|h-Nh| E\sum_{i\in\mathcal{J}} |E^{\{X_j\}_{j\notin S_i}} X_i|$$

$$+ \sup|h'| E\sum_{i\in\mathcal{J}} |X_i|(\sum_{j\in S_i} X_j)^2.$$

I have nothing to add to the description of this lecture given in the introductory paragraph. Some idea of possible applications of this approach can be obtained from my paper in the Proceedings of the Sixth Berkeley Symposium, to which I referred earlier.